environmental engineer

Focus on Environmental Background Data Analysis

By Dave Barclift, Jeff Heath and Andy Drucker

The Navy's Installation Restoration Program is looking to better focus investigations and response actions at past waste operation sites by using two distinct methods for analyzing background constituent levels. The two methods are the "comparative method" and the onsite background data evaluation method, hereafter referred to as the "onsite method".

Both the comparative and onsite methods can be used to effectively identify background data concentration ranges by applying various geochemical and/or statistical techniques. The comparative method has been used with varying success at a number of sites, while the onsite method is gaining regulatory acceptance and also should be considered when differentiating background conditions from releases.

The major challenge for both methods is to differentiate releases of site operation related chemicals from background conditions. A high variability or heterogeneity in the composition of soils makes it difficult to establish a single universal background concentration for soils or sediments in a region. EPA 1995 guidance indicates that it is more useful to discuss the range of background concentrations for a contaminant than to identify a single background concentration.

Two types of background conditions are considered when evaluating environmental background conditions. They are natural background and anthropogenic background. Natural background is defined as concentration ranges of naturally occurring constituents that have not been influenced by human activities. Examples of natural background constituents include metals derived from native rocks, asbestos fibers, hydrocarbons from oil seeps, and polynuclear aromatic hydrocarbons (PAHs) from forest fires.

Dave Barclift is with the Naval Facilities Engineering Command in Lester, Penn. Jeff Heath and Andy Drucker are with the Naval Facilities Engineering Command in Port Hueneme, Calif.

Anthropogenic background is defined as widely distributed chemicals present in the environment because of human activities, but not related to an onsite release. Examples of anthropogenic background constituents include atmospheric deposition of lead from the offsite use of leaded motor vehicle fuels and contaminating onsite sediments by offsite agricultural storm water runoff.

Under the Navy's Installation Restoration Program, there is no requirement to clean up background constituents at past waste sites. Therefore, it is important to draw a distinction between background conditions and contaminants of potential concern (COPCs) that can be attributed to releases associated with past and/or present site activities.

Comparative Method

The comparative method is the more conventional approach for identifying COPCs. Sample data collected from a nearby uncontaminated or "background" site are compared to data from samples collected at a site suspected to be contaminated. A variety of statistical comparison tests are performed using both data sets to determine if the ranges are statistically similar. These

Substance	Number of Sample Results	Degrees of Freedom	Statistical Distribution of Data
Arsenic-MK	10	9	lognormal better fit than normal
Arsenic-NE	10	9	normal better fit than lognormal

Table 1. Statistical distribution of subsurface soil data for mucky peat (MP) and silty loam (SL) for U.S. East Coast landfill.

comparative tests are used to identify background constituents, but they cannot be used to establish a background concentration range. However, if COPCs are identified using comparative statistical testing, a background concentration range can be determined by evaluating probability plots or by employing a threshold calculation method (e.g., 95th % upper confidence limit).

Onsite Method

The onsite method is based solely on data from the suspected contaminated site to extract background concentration ranges. No offsite data are needed. The onsite method uses geochemical and statistical methods to extract background data from onsite data. This innovative method consists of three possible steps.

- First, semi-qualitative data analysis and probability plotting is performed on the data at various locations and sample depths to evaluate trends. The result of this analysis may distinguish background levels of a contaminant from "outlier" levels that may be indicative of a site release.
- If outliers are observed, the next step is to investigate geochemical correlation between naturally occurring metals and a COPC at the site. As a result of geological and geochemical processes, certain elements, i.e., metals, are associated with each other. For example, data analysis may find a statistical correlation between the COPC, for example, chromium, and a non-COPC, for example, aluminum. A strong correlation between the metals is indicative of naturally occurring metals concentrations.
- If this analysis is inconclusive, the third step uses geochemical enrichment factor analysis to further evaluate natural geochemical enrichments. During geochemical processes, many metals are enriched in the soil, e.g., due to adsorption of metals onto clays, and these natural enrichments can be distinguished from enrichment due to site operations, i.e., a release.

The onsite method has the advantage of reduced sampling and analysis costs, as there is no need for offsite background data. Additionally, the onsite

Results of Shapiro-Wilk or Shapiro-Francia Distribution Tests			Standard Deviation or Log Standard Deviation	Arithmetic Mean of All Site Results	Maximum Positive Site Concentration
W-norm.	W-lognorm.	W-Table			
0.8786	0.9345	0.842	0.397	6.45	12.3
0.9709	0.9561	0.842	2.53	6.24	10.8

Continues on page 12 →

Substance	Surface Soil Detection	Subsurface Detection	Mean or Log of Mean	Standard Deviation or Log	t value	Surface Soil UTL	Subsurface Maximum	Conclusion Subsurface
Arsenic - MK	10/10	1/1	1.79	0.397	1.833	12.9	3.2	N
Arsenic - NE	10/10	1/1	6.24	2.53	1.833	11.1	5.7	N

Table 2. Arsenic Upper Tolerance Limit (UTL) for MP and SL Soils, Subsurface vs. Surface Soil UTL, for U.S. East Coast landfill.

Substance	Number of Sample Results	Degrees of Freedom	Statistical Distribution of Data	Results of Shapiro-Wilk or Shapiro-Francia Distribution Tests			Standard Deviation or Log Standard Deviation	Arithmetic Mean of All Site Results	Maximum Positive Site Concentration
				W-norm.	W-lognorm.	W-Table			
Arsenic-MK	11	10	lognormal better fit than normal	0.8837	0.9513	0.85	0.422	6.15	12.3
Arsenic-NE	11	10	normal better fit than lognormal	0.96	0.9471	0.85	2.41	6.19	10.8

Table 3. Statistical Distribution of Combined Soil Data for Mucky Peat (MP) and Silty Loam (SL) for U.S. East Coast landfill.

Concentration Range of Postive Results	MK Background Soil data Distribution	NE Background Soil Data Distribution
No. of samples	3	3
25% Quantile	3.9	4.4
No. of samples	6	6
75% Quantile	8.4	7.4
No. of samples	2	2
95% Quantile	12.3	10.8
Maximum Concentration	12.3	10.8

Table 4. Quantile Range Distributions of Background Soil Data for MP and SL soils U.S. East Coast landfill.

Concentration Range of Postive Results	MK Background Soil data Distribution	NE Background Soil Data
Frequency of Detection	11/11	11/11
Minimum Detected Concentration	3.2	2.1
Minimum Qualifier	J	J
Maximum Detected Concentration	12.3	10.8
Maximum Qualifier	J	J
Mean of All Data	6.15	6.19

Table 5. Occurrence and Distribution of Arsenic in MP and SL Soils at U.S. East Coast landfill.

method avoids the situation where the offsite background site is not geologically and geochemically similar to the site.

The following two case studies illustrate how each method can be applied to determine natural background concentration ranges.

Comparative Method Case Study

The comparative method has been used at a number of Navy sites. The case study below demonstrates how this background analysis method was successfully applied.

Site History

From World War II until 1955, a Navy landfill located on the U.S. East Coast received an unknown quantity of various industrial wastes. In the 1980's, the site was excessed to the State, and later sold to a private concern, its present owner. Under the property transfer agreement, the Navy is responsible for environmental restoration.

In 1997, the Navy completed a site investigation in accordance with the State's remediation regulations. Based on the findings and planned marina use by its present owner, the State required remediation to residential standards. The site contaminants exceeding these standards are arsenic, lead, and total petroleum hydrocarbons. The State residential direct exposure criteria soil objective for arsenic above groundwater elevation is 1.7 mg/kg.

In accordance with the State's remediation regulations, the Navy conducted a site-specific background soil investigation. The term "background" as defined in the State regulations includes both non-anthropogenic and anthropogenic constituents. The investigation was to determine whether soil in the vicinity of the landfill site contains naturally occurring arsenic at levels higher than the 1.7 mg/kg soil objective and to determine an appropriate arsenic soil remediation level for the landfill.

Data Collection

The landfill background soil investigation included a historic land-use review; a reconnaissance survey to evaluate proposed background sampling locations, soil sampling and analysis of 20 background soil locations, plus quality assurance/quality control samples; and a global positioning survey of the background sample locations. The reconnaissance survey was conducted with the State project manager and a U.S. Soil Conservation Service soil scientist who assisted in soil type identification.

Based on soil survey maps presented in the United States Department of Agriculture soil survey, two soil types were identified as representing soils at the landfill site prior to landfilling activities. The soil types found to be representative were mucky peat (MP) and silty loam (SL). Ten locations were sampled where each soil type was present in areas not impacted by the landfill.

Statistical analyses were performed following completion of the background soil sample analyses. The arsenic background analytical data underwent several statistical comparisons to determine whether it was appropriate to treat data from the different soil types separately or to combine them into a single background data set. All statistical analyses were performed in accordance with the guidance and recommendations presented in Navy, EPA, and related references.

Data Evaluation

The first step of the data evaluation was to determine whether the surface soil data were suitable to combine with the subsurface soil data.

For both MP and SL soil, ten surface soil and one subsurface soil samples were collected. Table 1 presents W-Test results evaluating the distributional shape for arsenic. Based on best fit, the distribution for the MP soil was assumed to be lognormal in shape, however, a normal assumption could not be rejected. For the SL soil, the W-Test showed that a normal distribution was the best fit, although lognormal assumption could not be rejected. Assuming a lognormal distribution for MP and a normal distribution for SL, a 95% upper tolerance limit (UTL) was estimated for the surface soil data set. Table 2 shows that the subsurface arsenic concentration does not exceed the respective surface soil 95% UTLs. No other statistical tests were performed because of the small number of subsurface soil samples. Based on the UTL test results, the assumption that the surface and subsurface data sets are from statistically equivalent populations was not rejected, so these data sets were combined for both the MP and SL soils.

The next step includes evaluating whether the MP and SL soils data sets have statistically similar distributional shapes. Exploratory data comparisons determined an acceptable degree of similarity in the distributional shape of MP arsenic data versus SL arsenic data. Table 3 presents the results of the W-test evaluating the distribution of MP, combined surface and subsurface soil. Again, based on best fit, a lognormal distribution was assumed. Table 3 also presents the results of the W-test evaluating the distribution of SL, combined surface and subsurface soil. Based on best fit, a normal distribution was assumed.

Descriptive statistics also revealed a similarity between

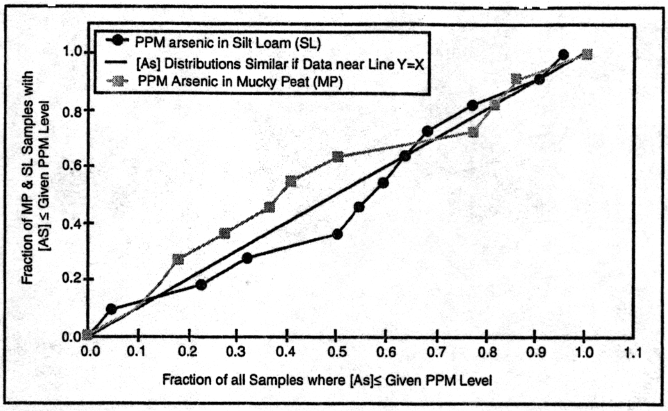

Figure 1. Comparison of frequency distributions for arsenic in two soil types: mucky peat (MP) vs. silty loam (SL).

Frequency of Detection	Log of Mean	Log of Standard Deviation	t value	UTL
22/22	1.74	0.413	1.7207	11.8

Table 6. Statistical Distribution of Combined (MP + SL) Background Soil Data at U.S. East Coast landfill.

these two soil types. Table 4 shows that the two data sets have similar arsenic concentrations at the 25th, 50th, 75th, and 95th quantiles. Table 5 displays the close similarity of the maximum, minimum, and average arsenic concentrations between the two data sets.

Figure 1 shows that the cumulative frequency plots of the two data sets are also similar based on a nearly linear match of the cumulative frequency distribution from each soil type to the cumulative frequency distribution for the combined data set.

The third step involves a series of comparative statistical analysis. These analyses were performed to determine if either data set contains arsenic concentrations that are statistically greater than arsenic concentrations in the other data set. The background data sets, MP and SL, were compared using:

- Student's t-test
- Satterthwaite t-test
- Bartlett's test
- Mann-Whitney test
- Gehan's test
- Quantile test
- Test of proportions
- Fisher's exact test

These tests demonstrate that the MP data set does not contain arsenic levels statistically greater than the SL data set nor does the SL data set contain arsenic levels statistically greater than the MP data set.

Therefore, the data comparison steps indicate that the two databases can be combined. Now that the data steps are combined, we can move on to the fourth step which includes distributional analysis (Table 6), descriptive statistics (Tables 7 and 8), and UTL

Continues on page 14 →

calculation (Table 9) as was done previously for the individual data sets.

The 95% UTL for arsenic in the background data set was estimated at 11.8 mg/kg, based on the assumptions of a lognormal background population. The 95% UTL is defined as a tolerance limit expected to contain 95% of all possible measurements for the background data set. Inherently, use of the 95% UTL test on several site-related samples can lead to a high false positive rate according to Navy guidance; therefore, exceedances should be confirmed by additional population comparison.

Summary

Using this comparative statistical approach, the Navy and the State were able to develop a site-specific background concentration for arsenic. This concentration was then used as a remediation standard for arsenic in soil, saving the Navy approximately three million dollars.

Onsite Method Case Study

The onsite method has been used at a number of Navy sites. The case study below demonstrates how this background analysis method was successfully applied.

Site History

Periodically in the past, the study site and surrounding area was reportedly used by several Federal agencies as a disposal site. Also, from the 1940's through the 1970's, the central western portion of the site served as an encampment for construction workers and a military hospital. A vehicle maintenance shop was also operated in the vicinity of the site.

The discovery of stained soil and buried scrap metal during construction led to a site investigation to assess the nature and extent of contamination. Soil gas, surface soil, and subsurface soil samples were collected throughout the adjacent site. Elevated levels of volatile organic compounds (VOCs), semivolatile organic compounds (SVOCs), (PAHs), metals, explosive residues, and TPH were detected in a pile of scrap metal. The scrap pile was cleaned up in accordance with State and Federal requirements. The site investigation report recommended that the Navy investigate the possibility of contamination extending onto adjoining property.

A remedial investigation (RI) was performed to characterize potential contamination suspected to be present based on the previous site investigation. Objectives of the RI were to:

- Determine if the site contains detectable levels

Continues on page 16 →

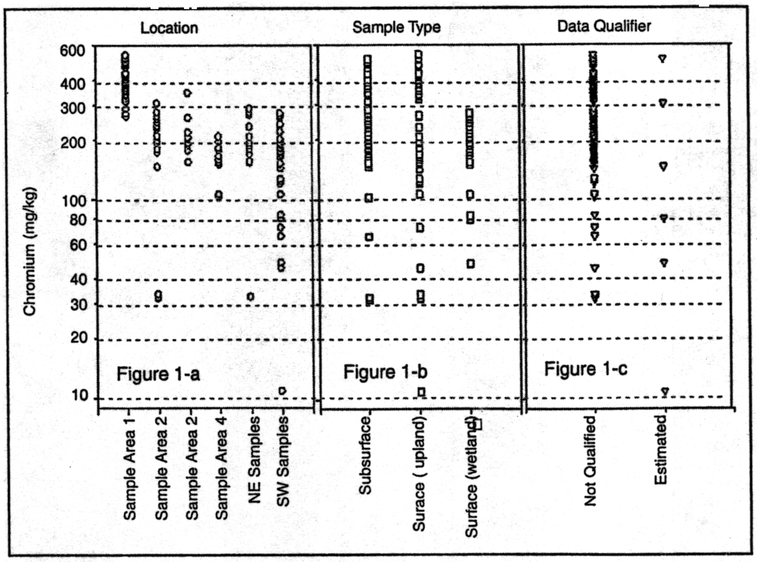

Figure 2. Semi-qualitative analysis.

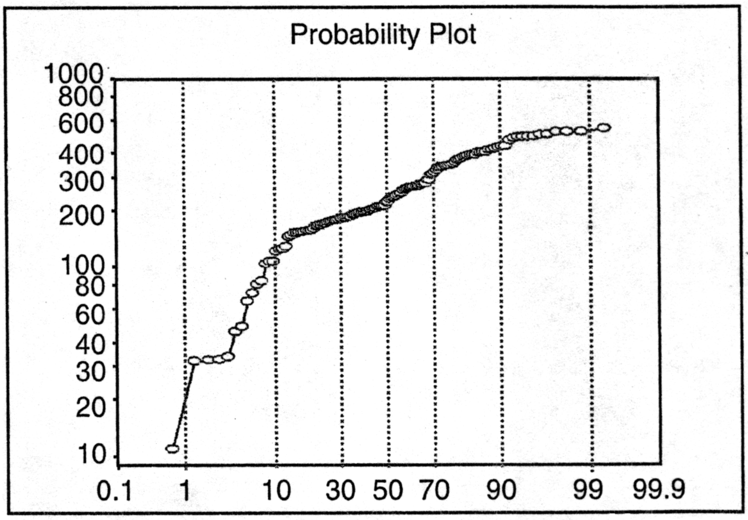

Figure 3. Cumulative probability analysis.

Concentration Range of Positive Results	MK + NE Background Sell Data Distribution
Number of samples	5
25% Quantile	4.25
Number of samples	12
75% Quantile	7.65
Number of samples	4
95% Quantile	12.1
Number of samples	1
Maimum Concentration	12.3

Table 7. Occurrence and distribution of arsenic in combined (MP +SL) background soil data set at U.S. East Coast landfill.

Concentration Range of Positive Results	MK + NE Background Sell Data Distribution
Frequency of Detection	22/22
Minimum Detected Concentration	2.1
Minimum Qualifier	J
Maximum Detected Concentration	12.3
Maximum Qualifier	J
Mean of All Data	6.17

Table 8. Quantile range distributions of combined (MP + SL) background soil data at U.S. East Coast landfill.

Substance	Number of Sample Results	Degrees of Freedom	Statistical Distribution of Data	Results of Shapiro-Wilk or Shapiro-Francia Distribution Tests			Standard Deviation or Log Standard Deviation	Arithmetic Mean of All Site Results	Maximum Positive Site Concentration
				W-norm.	W-lognorm.	W-Table			
Arsenic- MK + NE	11	10	lognormal better fit than normal	0.8837	0.9513	0.85	0.422	6.15	12.3

Table 9. Arsenic Upper Tolerance Limit (UTL) based on combined (MP + SL) data Set.

of contamination;

- Characterize the nature and extent of contamination resulting from past disposal and burial practices; and
- Determine the potential risk to human health and the ecology from possible site contamination.

The site evaluation concluded that VOCs, SVOCs, PCBs, TPH and explosives are not COPCs for the site because they were not detected or were below default screening criteria. However, a few metals including arsenic, chromium, and lead were identified as COPCs because they were detected in surface soils at elevated concentrations above default screening levels. The following presentation uses chromium to illustrate the onsite method.

Data Evaluation

To evaluate the elevated chromium concentrations the onsite method was employed. The first two steps of the onsite method are: (1) semi-qualitative analysis and probability plotting, and (2) evaluation of geochemical correlation.

As part of the semi-qualitative data analysis, chromium concentrations were plotted as a function of sample location, sample depth, and data qualifier (Figure 2). The plots in this example demonstrate the distribution range of chromium concentrations in soil for the site. Sample concentrations were plotted on a Log_{10} scale. Using the log distribution graphs, the difference and range between significantly high, i.e., potentially contaminated, and normal, i.e., natural background, metal concentrations can be observed visually. The three different semi-qualitative graphs shown in Figure 2 include:

- Concentration versus location. This graph shows the range of chromium concentrations at each site location (Figure 2-a). The concentration versus location plot can indicate potential outliers that may be representative of a release. For example, the high chromium concentrations detected mainly in Sample Area 1 may represent a release that should be further evaluated.
- Concentration versus sample type. This graph shows the range of chromium concentrations for subsurface, surface upland, and surface wetland sampling locations. All soil samples collected from 0 to 0.5 feet below ground surface (bgs) were considered to be surface samples; all soil samples collected below 0.5 feet bgs were considered to be subsurface samples. All samples collected outside of the wetland boundaries were classified as upland samples. As demonstrated in Figure 2-b, the chromium concentration ranges are almost identical between surface upland and subsurface soil at the site. This may be indicative of natural conditions, as releases usually demonstrate a concentration gradient between surface and subsurface sampling locations.
- Concentration versus data qualifier. This graph shows chromium concentrations versus data qualifiers assigned as a result of the data validation process (Figure 2-c). It is important to ensure that good quality data are used during the background data evaluation process. Conclusions made by using censored or non-detect data may lead to erroneous conclusions.

A probability plot was also prepared to further evaluate the distribution of site metals results (Figure 3). This plot was prepared by plotting the chromium cumulative percentage on a probability scale versus chromium concentration on a common log scale. This plot was then reviewed to evaluate distinct changes in slope, i.e., inflection points. The upper inflection point on the resultant curve may be indicative of outliers or outlier ranges from an assumed lognormally distributed population of analyte concentrations. No distinct changes in slope were observed within the

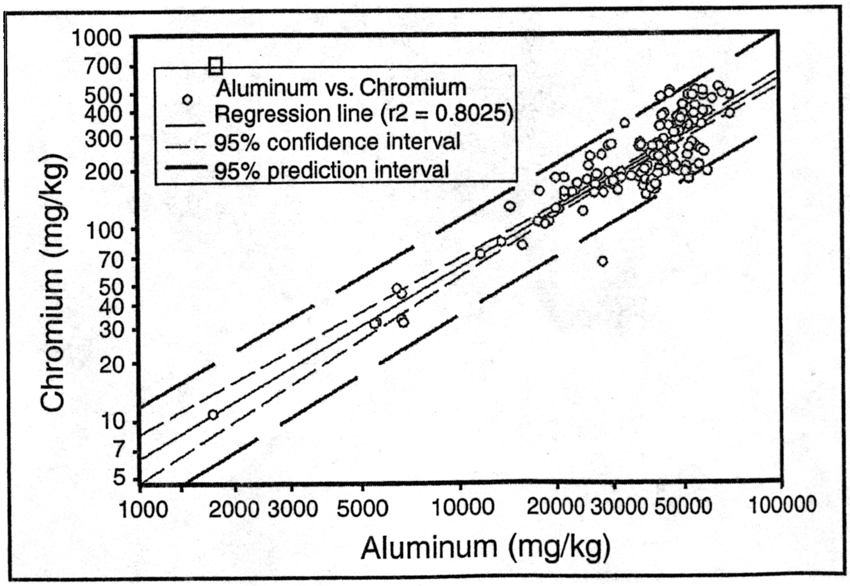

Figure 4. Geochemical correlation analysis.

upper 50% chromium population shown in Figure 3. Although not totally conclusive in this case, the probability plot suggests only one chromium population is present, i.e., it is representative of background conditions. While the probability plot suggests the chromium concentrations are representative of natural background conditions, i.e., presence of only one population, the high chromium concentrations detected mainly within Sample Area 1 (Figure 2-a) required further evaluation.

Geochemical correlation analysis was then used to confirm the results of the semi-qualitative data analysis and probability plot analysis. Geochemical correlation measures the strength of the linear relationship (r) between various metals. According to geochemical and geological processes, certain metals are correlated or associated with one another in predictable ratios indicative of the parent rock materials from which the soils were derived. If a COPC (metal) is strongly correlated with a natural occurring metal, the COPC should not be considered a COPC and should be considered to be natural background.

For this investigation, the correlation between chromium and aluminum was evaluated by plotting their respective concentrations on a log to log scatter plot (Figure 4). Aluminum was used due to its high natural abundance in site soil, its relatively small inputs from anthropogenic sources, and its normally distributed population. Other naturally occurring metals such as calcium, iron, magnesium, and silicon can also be used to conduct geochemical correlation.

To model the geochemical correlation, the least-square regression line (r^2) is typically used. This regression line consists of a straight line that describes how a response variable like chromium changes in response to an explanatory variable like aluminum. As demonstrated in Figure 4, the regression line (r^2 =0.8025) explains 80.25% of variation in chromium concentration with a 95% confidence level. All but one of the observed chromium concentrations are within the upper and lower limits of the 95% prediction interval, i.e., shows very good correlation. More importantly, high concentrations of chromium in Sample Area 1 are strongly correlated with high naturally occurring concentrations of aluminum. Based upon this, the chromium detected at the site was considered to be indicative of natural or background concentrations. As a result, chromium was eliminated form the list of COPCs for the site.

To access background data analysis guidance documentation on the comparative method, please visit the Naval Facilities Engineering Service Center's web site at http://erb.nfesc.navy.mil/restoration/methodologies/main.html . ∎